DOCUMENT RESUME

ED 454 860 IR 058 151

AUTHOR	Mann, Thomas
TITLE	Is Precoordination Unnecessary in LCSH? Are Web Sites More Important To Catalog Than Books? A Reference Librarian's Thoughts on the Future of Bibliographic Control.
PUB DATE	2000-11-00
NOTE	48p.; In: Bicentennial Conference on Bibliographic Control for the New Millennium: Confronting the Challenges of Networked Resources and the Web (Washington, DC, November 15-17, 2000); see IR 058 144.
AVAILABLE FROM	For full text: http://lcweb.loc.gov/catdir/bibcontrol/mann_paper.html.
PUB TYPE	Opinion Papers (120) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Cataloging; *Coordinate Indexes; *Indexing; Information Retrieval; Library of Congress Classification; Library Services; *Online Catalogs; Reference Services; *Subject Index Terms; *World Wide Web
IDENTIFIERS	*Library of Congress Subject Headings; Web Sites

ABSTRACT

Precoordination of Library of Congress Subject Headings (LCSH), in both the LCSH thesaurus and online public access catalog (OPAC) browse displays, continues to be necessary for several reasons: (1) the meaning of thousands of LCSH headings depends on their word order in ways that cannot be captured by postcoordinate Boolean combinations or by word proximity searches that drop relational prepositions as stop words; (2) a vast network of linkages between LCSH headings and the Library of Congress classification scheme depends on precoordination; (3) displays of precoordinated strings enable researchers to simply recognize whole array of relevant research options that they could never specify in advance in postcoordinate combinations; and (4) the precoordination of terms is inseparably linked to a vast network of cross-references that would vanish without it. In the future, LCSH must serve in both the environments of online library catalogs and the World Wide Web, not the latter in place of the former. An Online Cataloging-in-Publication (OCIP) program would enable the library profession to maintain the necessary precoordination of LCSH headings in OPACs and also to insert librarian-created LCSH elements into the Web headers of participating online publishers. This would enable users to exploit the existing precoordination and postcoordination capacities of OPACs and also to exploit LCSH more extensively in the exclusively postcoordinate search environment of the Web. (Contains 25 references.) (MES)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 454 860

Is Precoordination Unnecessary in LCSH? Are Web Sites More Important to Catalog than Books?
A Reference Librarian's Thoughts on the Future of Bibliographic Control

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Thomas Mann

Reference Librarian
Library of Congress
tman@loc.gov

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
B. Wiggins
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Summary

Precoordination of LCSH subject headings, both (partially) in the LCSH thesaurus and (more extensively) in OPAC browse displays, continues to be necessary for several reasons:

- The meaning of thousands of LCSH headings depends on their word order in ways that cannot be captured by postcoordinate Boolean combinations *or* by word proximity searches that drop relational prepositions as stop words.
- A vast network of linkages between LCSH headings and the LCC classification scheme depends on precoordination–i.e., changes in the word order of the subject strings also changes the classification areas to which the terms point.
- Displays of precoordinated strings enable researchers to simply *recognize* whole arrays of relevant research options that they could never specify in advance in postcoordinate combinations. The larger the file, the more such recognition capabilities are necessary.
- The precoordination of terms is inseparably linked to a vast network of cross-references that would vanish without it.

Books are not vanishing or generally evolving into digital forms; they continue to be published in huge numbers every year, and they provide formats that are more readable for lengthy texts.

In the future, LCSH must serve in both the environments of online library catalogs and the Web–not the latter in place of the former.

An Online CIP (OCIP) program would enable our profession to maintain the necessary precoordination of LCSH headings in OPACs *and also* to insert librarian-created LCSH elements into the Web headers of participating online publishers. This would enable us to exploit the existing precoordination and postcoordination capacities of OPACs, and also to exploit LCSH more extensively in the exclusively postcoordinate search environment of the Web.

IR 058 151

BEST COPY AVAILABLE

LCSH headings in copy cataloging cannot be simply accepted "with little or no modification."

Is Precoordination Unnecessary in LCSH? Are Web Sites More Important to Catalog than Books?
A Reference Librarian's Thoughts on the Future of Bibliographic Control

Thomas Mann

Aristotle wrote that "The least initial deviation from the truth is multiplied later a thousandfold"; Mortimer Adler similarly paraphrases Thomas Aquinas in saying "little errors in the beginning lead to serious consequences in the end."[1] The point here is that participants in this Conference need to pay particular attention to initial, unargued assumptions about the very purposes of cataloging and metadata if we wish to ward off some very large unintended, but nonetheless very undesirable, consequences if those purposes are inadequately assessed right at the beginning.

My major concern is this: Some of the papers before this Conference suggest that the Library of Congress Subject Headings system (LCSH) can be tailored to the task of Web cataloging by eliminating–or at least substantially reducing–its precoordinate displays of subject strings, both within the basic list itself and within browse displays in online catalogs. There is even a suggestion that such browse displays of strings of terms are entirely unnecessary, given the computer's ability to do postcoordinate Boolean combinations. I will demonstrate in some detail that this belief–often apparently more assumed than forthrightly stated–is extraordinarily naive. If, as a result of this Conference, the researchers of this country lose precoordinated displays of terms in LCSH–which serve several definite functions that are apparently being overlooked–then future scholars will have much less efficient subject access to large book collections. The gains–if they come about–achieved in better access to Web sites will be more than vitiated if they are accomplished at the expense of losing access to large (and still growing) book collections by undercutting the many functions of LCSH that require precoordination.

One immediate recommendation

Before examining what I think are bad ideas, let me jump ahead to one recommendation that I hope this conference will consider. As a reference librarian I'd very much like to see browse displays like this in catalogs of the future, integrating references to both books and Web sites:

Women–Services for
Women–Services for–Bolivia–Directories
Women–Services for–Caribbean area–Case studies
Women–Service for–Ethiopia–Congresses
Women–Services for–Germany–History

Women–Services for–Michigan–Evaluation
Women–Services for–New Zealand–Bibliography
Women–Services for–North Carolina–Finance
Women–Services for–Study and teaching–United States
Women–Services for–Study and teaching–United States–Web sites (.edu)
Women–Services for–United States–Directories
Women–Services for–United States–Web sites (.com)
Women–Services for–United States–Web sites (.edu)
Women–Services for–United States–Web sites (.edu)–Data archives
[This "Data archives" subdivision may not be appropriate for this particular subject; I offer it here just as a pattern example.]
Women–Services for–United States–Web sites (.edu)–Discussion lists
Women–Services for–United States–Web sites (.edu)–Portals
[I'm using "–Portals" here; "–Site directories" might be an alternative, in which case a cross-reference is needed: Site directories USE Portals]
Women–Services for–United States–Web sites (.gov)
Women–Services for–United States–Web sites (.org)
Women–Services for–Wisconsin–Periodicals
Women–Services for–Zambia–Directories

Such a display would enable researchers to *recognize* selected, high quality Web sites *in relationship to* the substantive knowledge records in the library's book collections–which are not, and for the most part never will be, digitized. (Of course there should be live links from the catalog records to the Web sites insofar as licensing agreements allow.)

In contrast, reliance on exclusively postcoordinate combinations such as Women AND Services AND "Web sites" would conceal the *relationship* of the Web resources to the relevant books.

Both precoordination and postcoordination necessary

The presence of a such a precoordinated browse display, of course, does not preclude postcoordinate Boolean search capabilities. Neither I nor anyone else is arguing for precoordination *rather than* postcoordination. We need *both* browse displays of precoordinated strings *and* the possibility of postcoordinate combinations of individual elements.

Browse displays, above all, enable us to recognize search options that we could never specify in advance, in Boolean combinations, by showing them to us in relation to options that we can think of. The larger the file, the more researchers (and reference librarians) need this recognition capability. What I am afraid of is the dismissal, on inadequate grounds, of the continuing importance of browse displays of ordered subject strings.

The loss of precoordination in LCSH in the Web/networked environment would cause very serious retrieval problems if the same loss were extended to LCSH in the OPAC environment. Since there's no point in maintaining two different LCSH systems, these very real problems in the OPAC environment have to serve as a brake on the otherwise free-floating speculations, untied to real library collections, that inform many of the projections of LCSH's future when considered exclusively in the Web environment.

When speaking of precoordination in LCSH, we must distinguish two different locales in which subject phrases must be displayed, although to varying degrees: first, within the LCSH list itself; and second, within online catalog browse displays, which show the linkage of free-floating subdivisions to headings, not displayed in the list itself.

Meanings of LCSH terms and links to LCC dependent on precoordinated word order

The first reason that precoordination must continue to be shown in the LCSH list itself lies in the need to capture intellectual meanings dependent on word order or prepositional relationships that are not captured by postcoordinate Boolean combination, or by simple word-proximity searching.

Moreover, such ordered combinations often entail specific links to the classification scheme. The order of the words in the headings changes the Library of Congress Classification (LCC) areas to which the headings are linked.

For example, the string **Philosophy–History** is spelled out precoordinately in the LCSH list even though "History" is elsewhere a free-floating subdivision. Why does the relationship of these terms need to be spelled out like this? and why does it then need to be precoordinated *in the LCSH list* rather than simply *in the catalog's browse display* of other subdivisions under "Philosophy"?

The phrase needs to be precoordinated to begin with because the order of the terms changes the meaning of the phrase: **Philosophy–History** is not the same thing as **History–Philosophy**. The phrases need to be combined *in the list* because additional information about the subjects must also be conveyed to both catalogers and catalog users: that *a change in the order of the terms also signifies a change in the classification areas appropriate to the different phrases*:

- **Philosophy-History** is explicitly linked to a major clustering of books on this subject in the B69-B4695 areas of the classified bookstacks.

- **History–Philosophy**, on the other hand, is explicitly linked to the D16.7-D16.9 areas of the stacks.

This explicit linkage of LCSH to the Library of Congress Classification scheme (LCC) permeates the length and breadth of the subject heading list. (This important fact is simply overlooked in some of the

papers before this Conference. It is perhaps noteworthy that the *Sears List of Subject Heading* is linked to the DDC system in just the same way.)

A postcoordinate combination of History AND Philosophy (in Voyager, entered as +history +philosophy in the keyword search mode) will, first, exceed the system's display limit of 10,000 records in my library's catalog. Second, the display of the 10,000 records that *are* retrieved will show in its first fifteen items–the ones that are most highly "relevance ranked"–classes numbers scattered among B, BD, DA, DT, GV, HC, HG, JN, ML, PA, QA, and Z. Not only does the meaning of the words change when their precoordinate ordering is lost; the specific areas of the bookstacks most closely associated with those different meanings are also concealed from a researcher's view.

If the two terms, **Philosophy** and **History**, are searched not as keywords but as subject terms confined to the controlled 6XX subject fields, their postcoordinate combination will *still* produce (in my library's catalog) a retrieval in excess of the 10,000 records that can be displayed; and the first twenty that do show up and have class numbers are scattered among AS, B, BH, BQ, CB, GV, H, HV, LA, PQ, and Q areas. The reader will be overwhelmed with "relevance ranked" junk, and will also be prevented from knowing which stack areas would be best to browse for full-text information.[2]

Even faceted elements must sometimes be displayed in precoordinated strings

Even if there is, quite properly and usefully, much faceting in LCSH so that the same subdivision can be applied to many headings, the *display* of *some* heading-subdivision combinations must still be shown in precoordinated manner in the basic list. This is because the order of the words is often tied to particular classification "cluster" areas. Another example is the heading **Women–Services for**, which in our catalog (including all further subdivisions) turns up 176 records, with noticeable clustering of the referenced books in three class areas, HV1442-1448, HQ1236.5-1240, and the HQ1740s.

A relevance-ranked keyword search of Women AND Services (in Voyager, +women +services), however, turns up and overwhelming 1,797 records (of which 1600 are books). Of the "most relevant" fifteen displayed first, only two records show up in any of these three clusters, and in two separate ones at that (i.e., one gets a sense only of individual items, not of important clusters). In other words, the "relevance" ranking completely erases from a searcher's perception the existence of such aggregates in the bookstacks–groups of related books, shelved together, that *are* brought to his attention via the precoordinated subject strings.

Additional linkages between LCSH strings and LCC show up in the catalog, not in the thesaurus

In this case, it is noteworthy that the **Women–Services for** heading is, within the LCSH list

itself, explicitly linked only to HV1442-HV1448–but in the library's actual catalog, a search under this string will bring up records that show definite clustering in the two additional areas just mentioned. In other words, the linkage of LCSH to the classification scheme is by no means simply a "one to one" connection. Its full complexity is discovered only by actually searching the precoordinated headings *in the actual catalog*, at which point the retrieval of records under the various subject terms may indicate yet other important clusterings associated with a particular string–which clustering areas are not formally indicated by LCSH-LCC links within the thesaurus itself.

This may sound sloppy to theorists who don't use actual catalogs and bookstacks very often; but my own experience is that the many linkages just *work*. The relationship of LCSH and LCC is partly specifiable in the LCSH list; but, in large part, the full extent of the interconnectedness of LCSH and LCC is discoverable only in the library catalog itself. This network of interconnections probably defies fully coherent *a priori* specification; but it nonetheless *functions* in the real world to direct readers from headings in the catalog to particular areas in the stacks. I sometimes think of New York City's underground as an analogy–the intertwinings of water lines, sewer tunnels, heating ducts, and electrical and optical conduits probably cannot be full determined on an *a priori* basis simply by looking at a blueprint or schematic (analogous to the LCSH list); one has to actually go down into a manhole to grasp fully what's wrapped around what (analogous to the full catalog). The larger point, however, is that we naively tamper with such myriad interconnections at our peril–and we certainly shouldn't embark on such a course by naively *overlooking the very existence of these linkages in the first place.*

Another analogy would be that of language: language does not fully reduce itself to neat rules that can be specified *a priori*. It develops on its own, in ways that defy logic. Just so is the relationship of all of the LCSH-ed records in a library catalog to all of the LCC-classified books in the stacks: the former definitely point to the latter, but logical rules spelled out beforehand are not always the best guide to the connection. Over the course of a century, the connections "just growed." To pretend that they are not there, however, and to simply ignore the continuing need for the catalog's precoordinated headings to point to particular "clustering areas" in the classified areas of the bookstacks, would be to do enormous harm to our nation's research libraries.

Additional examples of term meanings and links to LCC dependent on precoordination

Postcoordination of the terms, then–if relied on as the sole means of subject searching–utterly destroys not only the meanings of different subjects that contain the same words, but also the indexing of the class scheme that takes place when the subject terms are displayed in meaningful precoordinate relationship-strings. *A change in the order of the words also entails a change in the classification areas.* Other examples:

Indian women is not the same as Indian AND Women
Indian women–Mexico is linked to F1219.3.W6

Indian Women–North America is linked to E98.W8
Indian Women–South America is linked to F2230.1.W6[3]

Jewish women (linked to HQ1172) is not the same as Jewish AND Women

Women alcoholics (linked to HV5137) is not the same as Women AND Alcoholics

Women clergy (linked to BV676) is not the same as Women AND Clergy

For the sake of researchers who continue to use the bookstacks of major American libraries--and especially for the sake of the advanced academics in a wide variety of disciplines who are not represented at this Conference–we cannot naively overlook this extraordinary web of relationships linking these phrases (both in the LCSH list and in actual catalogs using LCSH) to the classification scheme.

A searcher who makes use of the precoordinated headings will thus be given important "focusing" information regarding which areas of the stacks to go to for the best groupings of knowledge records–*books*–for in-depth searching of *full-texts, back-of-the-book indexes, and prefaces* relevant to her topic–which knowledge elements are *not in the OPAC or on the Web*. A searcher who relies on postcoordination of separate elements will be overwhelmed with junk, and, further, will have no idea which stack areas would be best to examine first.[4]

Precoordination needed to capture prepositional relationships

Other terms need to be precoordinated in LCSH because *prepositional relationships* are crucial to the meaning of the terms–and prepositions vanish as stopwords in both postcoordinate Boolean combinations and word-proximity searches.

For example, searchers who browse **Women *on* television** will find 53 titles and be pointed, in LC's catalog, to particular *clusters* in PN1992.8.W65 and PN1995.9.W6. Searchers who browse "Women *in* television" will find the heading **Women in television broadcasting**, which will identify a third cluster of records at HD6073.T382 (Classes of labor. Women. Special industries or trades). Only one book–not a cluster–in this HD area shows up under "Women *on*" rather than "Women *in*" television in LC's catalog.

Researchers who simply use the keyword "relevance ranking" software will, in combining Women AND Television (in Voyager, +women +television) will be inundated with 804 records, only 345 of which are book records; and of the top twenty "relevance ranked" records (disregarding unavailable in-process or incomplete CIP records), *none* fall into any of these three most-relevant clusters in the bookstacks. The indexing function that the catalog serves in relation to the classification scheme is utterly lost without precoordination.

Once again, postcoordination of separate words effectively erases important information linked to the precoordinated term-order in the subject heading. From the existing browse displays of the ordered subject strings, however, researchers are effectively guided to go *here*, *here*, and *here* for the best groupings of in-depth (full text) knowledge records in the bookstacks. Without such direction to the stacks provided by precoordination in LCSH, researchers in this country will have a much more difficult time finding substantive knowledge records–books–in libraries.

Additional examples of prepositional relationships requiring precoordination

Other examples of prepositional relationships and indexing information that would be lost without precoordination:

Motion pictures for women (linked to PN1995.9.W6) is not the same as Motion pictures AND Women

Photography of women (linked to TR681.W6) is not the same as Photography AND Women

Sexual ethics for women is not the same as Sexual ethics AND Women

Social work with women is not the same as Social Work AND Women

Violence in women is not the same as Violence AND Women

Women, Black, in art is not the same as Women AND Black AND Art

Women in advertising is not the same as Women AND Advertising

Women in art (linked to N7629-N7639) is not the same as Women AND Art

Women in communication (linked to P96.W6) is not the same as either **Women–Communication** or Women AND Communication

Women in development (linked to HQ1240) is not the same as Women AND Development

Women in the Bible (linked to BS57.5) is not the same as Women AND Bible

Women in Church work is linked to BV4415

Church work with women is linked to BV 4445

Church work with women–Catholic Church is linked to BX2347.W6

If we do not maintain such precoordinated displays in LCSH and in catalog browse displays, this Conference will be seriously crippling the field of Women's studies–we will be making it much more difficult for scholars in this area not just to find, but to get a *structured overview* of books relevant to their topic within research libraries.

The Goal of Cataloging

Let's keep in mind that the goal of *cataloging* is not simply to give researchers "something." That goal can nowadays be accomplished by simple keyword searching without any intelligent human intervention in the forms of categorization, standardization of terminology, linkage of disparate concepts, and structured displays of search options. The goal of cataloging, in contrast, is to give researchers *an overview of the extent of the relevant resources* available for their topics (this is a year 2000 paraphrase of the intent of Cutter's "what the library has"). Overviews require connections, cross-references, and displays of options that cannot be specified in advance by researchers who literally don't know the fields they're getting into, and who often barely know how to phrase their initial questions. Overviews require displays of *relationships*, not just isolated data. These cannot be achieved without some measure of precoordination.

"Heavy lifting" capability required in research libraries

I realize that maintenance of precoordination makes LCSH more complex than it would be if it were simply an entirely faceted system of individual elements available for postcoordinate Boolean combinations or word-proximity searches. But complexity is sometimes simply necessary in order to get important jobs done. The control panel of a giant C5-A transport plane is necessarily much more complex than that of a Piper Cub twin-seater. If the Air Force were to reduce the former to the simplicity of the latter, they would soon find that their major "heavy lift" vehicle is capable of transporting materiel only by taxiing along the ground for short distances instead of flying with heavy loads over long distances. They would lose their ability to *lift* heavy loads into the air.

In a similar way, research libraries have to maintain their "heavy lifting" capacities with their unparalleled local resources, inside their walls. (It is especially the "heavy lifting" capacities that United States libraries have in providing *subject access* to their collections that make them the envy of other libraries–and scholars–throughout the rest of the world.) Granted, not every researcher needs the full capacities of the retrieval system for every inquiry. But the full capacities still have to be maintained for the frequent and unpredictable times when they are needed. To return to the plane analogy, our country doesn't need a C5-A every time a package needs to be delivered; but it does need the C5-A to be in readiness at a moment's notice.

Isn't the level of our intellectual research capacity–*which is our profession's*

responsibility–just as important to this country as its military capacity? My experience as a reference librarian is that even questions that initially sound very "simple," from ordinary citizens rather than advanced scholars, often have a way of quickly escalating into inquiries that do indeed require the "heavy lifting" capacities of libraries. Whenever that happens we must be able to respond with more than just "something." We need to be able to map our way efficiently into the range of knowledge records available, not just respond with isolated information.

If we as professionals are not making knowledge available–in its largest possible frameworks of relationships, interconnections, and linkages–rather than just isolated bits of information, then we are nothing at all. If we see ourselves as providing access only to information rather than knowledge, or to information as a higher priority than knowledge, then we can indeed be replaced by machines.

Effects on Women's studies and Black studies

If we throw away precoordination in LCSH–which gives us so much of our "heavy lifting" capacity–we will be crippling not just the field of Women's studies, but that of Black studies: the arrays of precoordinated headings starting with the term **Afro-American(s)**–apparently soon to be changed to **African American(s)**–is fully as complex as the array of **Women** headings. I urge everyone participating in this Conference to take a look at the red books' thirty-five columns of precoordinated **Afro-** phrase headings arrayed on twelve pages–and this even without free-floating subdivisions being fully displayed.

Several times I have helped students who came in saying that they had to write a paper on "Black history." By alerting them to the amazing bounty of options they never knew they had, spelled out for their simple recognition just within the LCSH list (let alone within the catalog's browse display), such students are enabled to focus their topics in a wide variety of ways that would simply not otherwise occur to them. **Afro-American healers**, **Afro-American pacifists**, **Afro-American outlaws**, **Afro-American orchestral musicians**, and **Afro-American whalers** are all part of Black history; and these are only a very tiny sampling of the hundreds of options that would simply vanish from the radar screen if the searchers tried only **Afro-Americans** AND **History**.

Giving researchers *overviews of what is available*–opening up their eyes to unsuspected possibilities, positioning them on conceptual maps of options, and anchoring them within relevant intellectual frameworks–*this* is what public service is about; it is not a matter of giving them simply "something."

The **Afro-** headings, too–just as with the **Women** headings–tie particular aspects of Black studies embodied in precoordinated phrases to widely different areas of the classification scheme. For a five-page example of this point–which I mercifully will not reproduce here–see my *Library Research Models* book (Oxford U. Press, 1993), pp. 33-37.[5] If we unwittingly destroy the precoordinated display of the **Afro-** headings we will simply decimate the research potential of Black studies scholars

in American libraries.

How would such a development be reported in *The Chronicle of Higher Education* or *Lingua Franca*? (What would Nicholson Baker have to say about it in *The New Yorker*?!) Would it reflect credit on us? Or would it show that, in order to remedy our whole profession's traditional inferiority complex, we sold subject access to book collections down the river in order to appear more "with it" in Web searching?–and did so with the full knowledge that, while librarians can reasonably structure access to book collections in research libraries, we will never be able to intelligently apply LCSH to more than a microscopic sampling of the *billions* of Internet sites available. Will it be reported that we gutted precisely the elements of LCSH that make it so useful *in* structuring access to book collections, in order to facilitate unstructured applications of individual terms (stripped of both their contextual strings and links to LCC) to Web site records? Why do we assume, in the first place, that anyone will turn to *library catalogs* for primary access to the Web when field is already taken by Google, AltaVista, NorthernLight, Hotbot and a dozen other more comprehensive search tools?

The Virtue of OPAC Coverage of Web Sites

If library catalogs are to cover Web sites–and indeed they should, selectively–then their virtue will be precisely in bringing Web sites *into relationships* with the *substantive knowledge records* that *books* are–especially since book collections, for copyright and preservation reasons alone, will always reside primarily off the Web, within library walls. We need to tie the two sources together, not sacrifice one to the other. And one part of the linkage of the two environments–another will be discussed below–will be brought about most effectively by extending rather than eliminating the range of our precoordinated browse displays in our catalogs, as in the **Women–Services for** example above.

Precoordinated Word Order Also Affects Cross-Reference Structure

There is yet another reason not to destroy the display of precoordinated strings in LCSH: not only does the meaning of subject terms change depending on the order of their words; not only does the huge web of linkages between LCSH and LCC depend on the word-order of the terms; not only do the meanings of proximate nouns in the same order need to be distinguished by different prepositional relationships–not only for all of these reasons does precoordination need to be maintained in the OPAC environment, but for another reason, too: the order of terms also *critically affects the cross-reference structure* between and among related terms. (Of course cross-references don't show up in Web-type searches–the software can't handle them. Does that mean that they're now *also* dispensable in the OPAC environment?) Let me give just two examples from the hundreds of thousands available:

The precoordinated phrase **Women-Psychology** (which is explicitly tied to HQ1206-HQ1216 in LCC) is linked by cross-references to:

RT Women–Mental health
NT Achievement motivation in women
Animus (Psychology)
Anxiety in women
Assertiveness in women
Body image in women
Cooperativeness in women
Helplessness (Psychology) in women
Leadership in women
Self-esteem in women
Self-perception in women[6]

This entire network of relationships would be lost if users could search only Women AND Psychology. Researchers could find only isolated information, not a web of knowledge relationships.

The precoordinated phrase **Afro-Americans–Education** (which is explicitly tied to LC2701-LC2853 in LCC) is linked by cross-references to:

BT Education–United States
RT School integration–United States
NT Afro-American students
Afro-American women–Education
Afro-Americans–Professional education
Afro-Americans–Scholarships, fellowships, etc.
Afro-Americans–Vocational education
English language–Study and teaching–Afro-American students
Segregation in education–United States
Segregation in higher education–United States

Once again, all of these displayed linkages that bring to researchers attention options they would not otherwise perceive–all would be lost if, in order to make LCSH more "flexible" for a Web environment, we throw away precoordination in the OPAC environment. (Do we really want to do this? As the kids these days say, Isn't this a "no brainer"?)

Key Functions of LCSH Being Overlooked

Unfortunately, none of these problems entailed by eliminating precoordination are even mentioned by key papers before this Conference. (Even beyond this meeting, there are many cataloging theorists out there who seem to think that the *only* function of precoordination is "to break up large files." Where do they acquire such blinders? Is this what is being taught in schools of library and information science? Perhaps less time in the academic ivory tower and more time working

behind public service desks in real libraries is indicated.)

Let me turn to several other assumptions that show up in some of the papers–all of which affect the precoordination/postcoordination issue–and that I think are "not ready for prime time."

Information and Knowledge Are Not the Same

The first of these is something to which I've already alluded. It is the assumption that *information* and *knowledge* are the same thing, and can be formally handled by retrieval systems in just the same way. I beg to differ.

First, there is a real hierarchy in the realm of human awareness. The lowest level is formed by *data*, the unorganized, unfiltered, and unevaluated raw material of thought, comparable to sense experience (although, I think, not reducible to it–but that's another paper). *Information* is at a higher level, reflecting an organization of data to the point that statements can be made about it, either true or false, and coherent or incoherent with other information. *Knowledge* reflects a still higher level of organization to the point that truth or falsity can be reasonably assured by tests of correspondence to, and coherence with, the world of experience and of other ideas; it requires that information be put into much larger frameworks of relation to the worlds of matter and ideas. This level includes discernment of patterns and *interconnectivities* within information, and the making of generalizations that are accessible to, and acceptable by, other people. (I won't belabor here the further levels of understanding and wisdom.)

Information simply does not have the degree of "truth-claim" upon us that knowledge has, because it does not have the *connectedness* and *relatedness* of knowledge; and, further, it also depends on all of the larger frameworks of knowledge, understanding, and wisdom for an assessment of its *worth*.

These are not merely academic distinctions; they have a material bearing on the very purposes, methods, and materials of *cataloging* and bibliographic control.

Conveying Knowledge Requires Larger Cataloging Structures and Linkages

Briefly: We ought not to be dismantling the larger structures and webs of *knowledge* that cataloging has created in order simply to achieve less costly access to *unintegrated information*. Access to *information* is much more amenable to automatic machine methods of indexing, without human structuring, than is access to *knowledge*; but automatic methods of gaining access to information are not sufficient to show researchers the *knowledge relationships embedded within LCSH subject-strings themselves, within their cross-references, and within their integral connections to the Library of Congress Classification (LCC) scheme.*

Screen Displays and Book Displays Change Readability

The next assumption that we need to examine is the assertion that knowledge is equally well conveyed by screen displays as by book formats. I doubt this very much. How many of us are now reading book length narrative or expository works–say, the equivalent of a 200-page book–on screen displays? I'm not talking about long lists of hits on Google or Yahoo, or long lists of directory information, or bibliographical listings, or long rosters in Ebay; I'm talking about long, coherent narrative or expository texts. Some are reading such things on screens, I'm sure; but I'll just remind everyone to examine his/her own reading habits before imposing theoretical projections upon everyone else. If we don't read long connected texts on screen displays ourselves, let's not force others to be shunted by our catalogs exclusively or even primarily to Web sites rather than printed books.

Knowledge–requiring longer attention spans to establish its connectedness–is much more readily conveyed by *book formats* than by screen displays of textual material, which most people recognize as being "slanted" to shorter attention spans.[7] If this is true–and I think it is–then this Conference should not cavalierly assume that future catalogs ought to be more concerned with Web sites than with books. Catalogs need to cover both–but not the former in preference to the latter. Let's not forget, right at the outset, that book formats are a proven medium for conveying knowledge, while the verdict on Web sites is truly not yet in–and may not be as rosy as some are assuming. (The additional problem of changing the focus of library catalogs from books to Web sites is that of *preservation*–it is neither inevitable nor even likely that electronic resources can be preserved at nearly the cost-efficiency of preserving books.)

I strongly agree with Walt Crawford and Michael Gorman's initial position in their book *Future Libraries: Dreams, Madness, and Reality*: "Let us state, as strongly as we can, that libraries **are not wholly or even primarily about information**. They are about the preservation, dissemination, and use of recorded knowledge in whatever form it may come . . . so that humankind may become more knowledgeable; through knowledge reach understanding; and, as an ultimate goal, achieve wisdom."[8]

The *book format* is by far the best vehicle that humanity has devised for conveying to itself the higher levels of knowledge and understanding, and the *research library* is the best vehicle that has ever been devised for making large collections of substantive knowledge records *freely* available, without prohibitive individual subscription costs or point-of-use charges, or on-the-spot printing charges. Most of the billion+ Web sites, of course, are not substantive; and a high percentage of those that are most desirable are generally confined by license agreements to particular terminals within walls, or to tightly-defined user groups–i.e., such sites *cannot* be tapped into freely by anyone, from anywhere, at any time. In that sense they are much like books: *freely* available only *within library walls*.

Library Catalogs Provide *Alternatives* to the Web

Library catalogs, if they are to have an important function in the age of Google, Altavista, and NorthernLight, would serve users best by directing them to selected, substantive sources of knowledge–especially to the abundance of sources that are not, and never will be, freely available to anyone, from anywhere on the Web. This means that catalogs will function best by presenting researchers not just with different ways to search the Web, but with substantive *alternatives to* the Web, especially copyrighted or licensed resources that cannot be found within the vast ranges of free Web sites. (Most users think of "the Web" as the free portions of it; I find this repeatedly when I show researchers our licensed databases–their question is always phrased as "Can I get this on the Web?," but their meaning is "Can I tap into this for free outside the library walls?")

Other Questionable Assumptions

Beyond the misleading assumption that information and knowledge are the same, there are other questionable assumptions that we need to be on our guard to spot, all of which may be found in current literature, and some of which show up in some of the papers before this Conference:

- that "knowledge" records, in general, are now making a "transition" to digital forms;
- that the *only* context in which we must regard the future of bibliographic control is one of shared Web access–i.e., that the context of continuously expanding and *localized book collections* need no longer concern us as a higher priority;
- that the functions of cataloging in the persisting *book collections context* can now be dispensed with–without even examining what those function are–insofar as they are not readily adaptable to the context of accessing Web sites;
- that, specifically, precoordination in displays of LCSH subject heading strings is no longer necessary either as (partially) enumerated in the LCSH list itself; or as (fully) displayed in "browse"screens in online catalogs, because postcoordination of individual elements renders such string-displays intellectually "unnecessary" or, worse, socially stigmatizes them as "old fashioned" (thereby precluding any objective assessment of their continuing functions)
- that researchers of the new millennium will choose *library catalogs*, to begin with, as their *primary* avenues of access to the Internet;
- that library catalogs, preeminently, must dominate the information landscape of the future by "seamlessly" leading researchers to *all* of the information they may need (rather than serving more modestly as one channel of access to *some*

important knowledge and information records).

- that catalogs will, ought to be, and *can* be used successfully–i.e., to give inquirers an overview of their research options and to lead them to the best information/knowledge on their subjects–by *untrained* researchers *in isolation*, that is, in the absence of any intervention by reference (or other) librarians, either beforehand in bibliographic instruction classes, or immediately at the point of use. (This would be analogous to Piper Cub pilots trying to fly C5-A transports, with their much more complex control panels, without any help.)

- that, rather than using catalogs to integrate the two contexts of knowledge records contained in books and substantive Web sites, catalogers of the future should markedly diminish their concern for books and concentrate on Web sites instead.

- that any concern for maintaining precoordination in LCSH should be dismissed *a priori* on the grounds that, because it first developed within manual catalogs, precoordination is a mark of outdated, "pre-high-tech" thinking. (This is nonsense. Precoordination makes *online catalogs* function much more efficiently.)

Are Books Evolving into Digital Forms?

Martin Dillon, in a (thankfully) "blunt statement," works from one initial assumption very different from my own:

> After a long and various evolution, knowledge representation settled into paper products for most of its output. Now we are shifting to digital forms for representing knowledge and to the Web as the primary distribution channel. This change will have profound consequences. There is little question, for example, that paper products will gradually be replaced by Web-accessible digital products.[9]

I respectfully beg to differ. Even F. W. Lancaster now has "Second Thoughts on the Paperless Society."[10] Walt Crawford, in his article "Paper Persists: Why Physical Library Collections Still Matter,"[11] makes a number of relevant points:

> What happens if the premises arguing for library conversion to digital fail? Logically, if the premises are invalid, then the conclusion is false or at least unsupported.
>
> * * *
>
> Reading from digital devices, whether portable or desktop, suffers in several areas–among

them light, resolution, speed, and impact on the reader–and there has been essentially no improvement in any of these areas in the last five years.

Many futurists have conceded this point. They now admit that people will print out anything longer than 500 words or so. It's just too hard to read from a computer, and it doesn't seem likely to get a lot easier. If every long text is printed out each time it is used, there are enormous economic and ecological disadvantages to the all-digital library: briefly, a typical public library would spend much more on printing and licenses than its current total budget and would use at least 50 times as much paper as at present.

Continuing Production of Book Formats in Huge Numbers

It is also worth noting that the new *Bowker Annual* (2000) has, just this year, radically revised upward its statistics on the number of books produced in this country in recent years; last year it recorded 1997 book title production as 65,769 titles; now it records 1997 production as 119,262 titles. Similarly the revision of the 1998 figure is from 56,129 to 120,244 titles. It seems more than questionable to assume that *books* are making "the transition" that is so cavalierly assumed in so much information science literature these days. Research libraries are still heavily anchored in print collections *as well as* in digital resources; and the latter simply are not the only context in which LCSH must function.

Significant Differences Between OPAC Cataloging and Web Metadata: Displays of Relationships

Mr. Dillon makes a further point, with which I do not disagree, in quoting a description of metadata:

> *Meta-information has two main functions:*
> - *to provide a means to discover that the data set exists and how it might be obtained or accessed; and*
> - *to document the content, quality, and features of a data set, indicating its fitness for use.*[12] [italics in original]

This is fine–as far as it goes. But *cataloging,* unlike metadata, has additional functions beyond these two, especially in the context of book collections. One such function that is of great help in public service work is:

- **to *relate* subjects to other "outside" topics both (a) through formal cross-references of BT, RT, and NT relations, and (b) through displays of alphabetically adjacent subjects whose connections to each other are not caught by formal cross-references.**

I have already exemplified point **(a)** previously. Point **(b)** may not be as familiar, so let me give an example of it: in LCSH **Monasteries** is linked to the *narrower term* **Monasteries, Coptic** *not by an NT reference*, but *simply by its alphabetical proximity.* **Monasteries** is similarly linked to the cross-reference **Monasteries, Cistercian USE Cistercian Monasteries**. And the alphabetical proximity of **Monasticism and religious orders** leads to *its* NT cross-references to **Child oblates, Clerks regular, Contemplative orders** and a host of other headings otherwise scattered imperceptibly throughout the alphabet. There are whole columns of headings related to **Monasteries**–which will lead researchers in many directions–that are not linked to each other by cross-references; but they are linked nonetheless by this other mechanism. A very brief display of only some of these contiguous related headings includes the following:

Monasteries
(linked to BX2460-BX2749 Catholic Church and NA4850 Architecture)
Monasteries, Armenian
Monasteries, Buddhist
Monasteries, Hindu
(linked to BL1243.72-BL1243.78)
Monasteries, Jaina
(linked to BL1378)
Monasteries, Syrian Orthodox
Monasteries and state
Monasteries in art
Monastery gardens
Monastic and religious life
(linked to BX2435)
BT Spiritual life–Christianity
RT Vows
SA *subdivision* Spiritual life *under names of individual religious orders*
NT Celibacy–Christianity
Eremetic life
Evangelical counsels
Retreats for members of religious orders
Spiritual direction
Superiors, religious
–History–Early Church, ca. 30-600
(linked to BX2465)
Monastic and religious life (Buddhism)
Monastic and religious life (Hinduism)
(linked to BL12266.85)
Monastic and religious life (Zen Buddhism)
Monastic and religious life in art

Monastic and religious life in literature
Monastic and religious life of women
(linked to BX4210-BX4216)
–Psychology
(linked to BV4205)
Monastic guest houses
USE Monasteries–Guest accomodations
Monastic libraries
(linked to Z675.M7)
Monastic profession
USE Profession (in religious orders, congregations, etc.)
Monasticism and religious orders
(BX385 Greek church)
(BX580-BX583 Russian church)
(BX2410-BX4560 Catholic church)

All of these displayed relationships and linkages–and scores more not listed here–would be lost without both precoordination and alphabetically-adjacent listing. Without the perceptible contiguity of **Monasteries** to these other headings, all of these paths to related knowledge records could never be noticed by researchers. (Nor, again, are they captured by the cross-referencing system of BT, RT, and NT.)

My experience in standing over researchers' shoulders and explaining LCSH to them is that very few people realize the extent, variety, and specificity of the terms available to them, without some such display enabling them to recognize the related terms they could never specify in advance via Boolean combinations. Researchers very much appreciate having these option-displays pointed out to them–they cannot think of them on their own.

Again, almost all of the alphabetically-adjacent related or narrower terms are themselves precoordinated phrases. Both their contiguity and their very existence, however, would vanish in a faceted LCSH system shackled exclusively to a postcoordinate search capability.

The Continuing Need for Reference Assistance, Over and Above Catalog Improvements, in the Total System

Doing research in large libraries is seldom "transparent" to users, even to those who limit themselves to the library's catalog; some instruction, either beforehand or at the point of use, is usually required. Without such guidance from reference librarians researchers routinely miss most of "what the library has"–let alone "what the Web has"–without realizing they've missed anything. Again, it's like Piper Cub pilots trying to fly C5-A transports; without some additional instruction, all they will be able to do on their own is taxi the larger plane along the ground–they won't be able to really exploit its

heavy lifting capabilities. (This is why I say catalogs alone cannot bear the burden of doing "everything" by themselves; in the operation of the total system, reference librarians are just as integral as catalogs and catalogers if the heavy lifting capability is not to be abandoned. And our culture requires the continuance of that capability.)

I think the present Conference would not be prudent if it were to assume, without any argument, that reducing the display potential of LCSH headings, dumbing down the complexity of the strings themselves, abandoning displays of their cross-reference connectsions, and severing their links to LCC, is the way to enable people to do *better* research: to exploit that "heavy-lifting" capacity needed in large libraries. We should indeed be aiming at that goal of promoting better research; but we should also realize that its accomplishment will necessarily entail many more factors than improving library catalogs alone. One such factor is providing reference help.

LCSH Unlike Other Thesauri

An additional fact that tends to be overlooked by anyone who would reduce LCSH to the shackles of faceted thesauri is that other controlled vocabularies deal almost exclusively with the literature of one topic area; LCSH, on the other hand, must deal not only with all possible subjects of knowledge–not just information–records, but with the endless relationships between and among them, in ways that elude simple Boolean and proximity searching. (Look again at the cross-reference, and alphabetical-adjacency, examples of **Women** and **Afro-Americans**.) Other thesauri, too, (save for the *Sears List* and its links to DDC) do not have to serve as subject indexes to classification systems for shelving full-texts in arrays that allow them to be quickly browsed down to the page and paragraph level.

Significant Differences Between OPAC Cataloging and Web Metadata: The Importance of Browse Displays of Precoordinated Strings

Yet another function of *cataloging* that shows up so often in the public service context is:

- **to relate the various aspects "within" one and the same subject to each other through browse displays of subdivisions within online library catalogs.**

Most of these subdivisions are "free floaters" and, like facets in other controlled vocabularies, are not displayed as linked to their parent term within the thesaurus itself. *The needed display, however, is accomplished elsewhere, within the catalog rather than within the thesaurus.*

In other words, to point out that many LCSH strings (i.e., those with free-floating subdivisions not recorded in the thesaurus) are not displayed precoordinately within the thesaurus itself is not an argument on behalf of saying, therefore, that *all* secondary terms in any string can be treated as "free

floating." This is literally a non-sequitur. Those free-floating subdivisions that are not displayed precoordinately in the list have two important characteristics: **a)** their ordering in relation to their heading is not needed to determine meaning, cross-referencing, or linkage to LCC; and **b)** their ordering in relation to their heading is indeed displayed precoordinately elsewhere, within OPAC browse displays. Even "faceted" free-floating subdivisions require precoordinated browse displays in OPACs–for without such recognition arrays, most researchers would never think of their existence in Boolean combinations. OPAC browse displays of contiguous subdivisions provide a structure that shows the extent of the subject's aspects–a structure that could never be guessed at by naive researchers entering unfamiliar subject territories.

For example, I have helped many readers who were interested in researching particular countries. One asked for help on the history of Yugoslavia. On his own he had tried a keyword search, but the Boolean combination he'd done of **Yugoslavia** AND **History** had overwhelmed him (and the computer system itself) with more than 10,000 records. So I showed him how to do a browse search that would bring up a full array of subdivisions under "Yugoslavia"; and of course this kind of display alerts the researcher to much more than the one subdivision "History." It also displays options such as:

Yugoslavia–Antiquities
Yugoslavia–Boundaries
Yugoslavia–Civilization
Yugoslavia–Description and travel
Yugoslavia–Economic conditions
Yugoslavia–Ethnic relations
Yugoslavia–Foreign relations
Yugoslavia–Intellectual life
Yugoslavia–Politics and government
Yugoslavia–Rural conditions
Yugoslavia–Social life and customs

I didn't stay to watch which aspects he chose; I just showed him how to scroll through the array. (He did get very excited when he saw "Antiquities" as an option, however.) The point is that all of these options might well be of interest to an historian of this (or any other country); but most researchers would never become aware of the *range of options* they have in researching such a topic without such a display. Further, several of these subdivisions are free-floaters not recorded in the LCSH thesaurus itself; but they do show up in the OPAC browse display. *All of these relevant paths would be lost*–and in fact *were* lost–in the reader's search for **Yugoslavia** AND **History** in a postcoordinate Boolean combination of separated facets.

Precoordinated Subdivision Strings Do Much More Than Just "Break Up Large Files"

The virtue of such precoordinate displays is not merely that they "break up large files" but that *they alert readers to whole areas of options relevant to their interests that they could not specify in advance.* Granted, if their only function were to "break up large files," then such break-ups could be done through postcoordination. But, contrary to the beliefs of some catalogers who evidently do not work with the public, this is *by no means the only function* of precoordinated subdivisions; and the "little error" of holding a mistaken assumption here will lead to "very serious consequences" for researchers who want not just "something" on their topic, but a structured overview of their research options. (I may not be articulating this very well, but the difference here is at least *like* the difference between information and knowledge–the levels relationship and interconnectivity are simply not the same.)

The Need for Recognition Capability When Prior Specification Cannot Work

One more (brief) example: I once helped a Classics professor who wanted to know how the Greeks would have transcribed animal sounds (e.g., quack, oink, meow). He was already familiar with the frogs' croaking recorded in Aristophanes' *The Frogs*; but he was interested in other animal sounds. The LCSH term **Animal sounds** looked promising, but wasn't; it just didn't work. (It did work, however, in the printed *Social Sciences and Humanities Index* to turn up an article on "Suetonius' Catalog of Animal Sounds"–a Latin list, apparently, that the professor said he would also pursue.) So I thought we might browse through the subdivisions under **Greek language** to see what might turn up. What did turn up was **Greek language–Onomatopoeic words**, which led to a dictionary that included animal sounds. (I don't read Greek myself, but the professor told me he was satisfied with the book.)

Now of course it could be said that a postcoordinate combination of **Greek language** AND **Onomatopoei?** would turn up the same result; and that would be a true statement. But it would also entirely miss the point: Who would ever think in advance to use "Onomatopoei?" as one of the elements *in* the combination? (Similarly, who would think beforehand of all the differently-phrased options under "Yugoslavia"?) The major virtue of precoodinated displays of subject strings is that *they bring to our attention options that we could never specify in advance.* And the larger the file that is being indexed/cataloged, the more necessary are such aids if the resultant retrieval is to be anything more than fragmentary and orphaned from relatives. Again, it's roughly the difference between finding *information* about a few isolated options you can specify, vs. gaining a *knowledgeable overview*–a map that shows both the existence and the relationships–of all of your options within the catalog. (Writers who rhapsodize about the wonderful ways of searching brought about by computers seldom mention how much *more* powerful the computer searches themselves become when they enable readers to see precoordinated strings in browse displays–displays that enlist the tremendous power of simple *recognition*.)

Catalogs Cannot Do Everything That Needs To Be Done

Let me also add that in the "Yugoslavia" case I also put under the reader's nose the wonderful current article on the country in *Europa Yearbook*, and the *Yugoslavia: A Country Study* (1992) volume from the old area handbook series. And I let him know that we could also easily find a variety of other concise overview articles from scores of specialized encyclopedias by using the *First Stop* and *Subject Encyclopedias* indexes (neither of which is computerized). There is no way on earth this man would have found these overview starting-points on his own by searching the library's catalog, especially with **Yugoslavia** AND **History**. Even if he'd seen the record for the area handbook volume–which does not have the word "History" anywhere on it–its special significance as a starting-point would not have leapt out at him.

Once again: the catalog alone simply cannot do *everything* that needs to be done for researchers; and this Conference should not be assuming that it needs to take on that function. F. W. Lancaster, in his "Second Thoughts on the Paperless Society" article[13], makes some cogent observations:

> The [library/information science] profession has greatly exaggerated the benefits of technology, especially in the area of subject access. Putting electronic databases in the hands of library users does not necessarily mean that they can be used effectively. . . . Merging several catalogs into one creates much larger databases that are even less useful for subject access than their components. . . . Unfortunately, the majority of librarians seem to assume that more access means better access. This is not necessarily true. For 30 years, studies have consistently shown that information services users really want access to the best information. They want tools or people capable of separating the wheat from the chaff. They want quality filtering.
>
> The profession seems to have lost sight of this. How else can one explain why so many librarians are head over heels in love with the Internet, a monster lacking a minimum of control of content? . . .
>
> The service ideal still exists to some extent in public libraries and school libraries. However, the more specialized the library becomes in the academic world, encouraging remote use, the more dehumanized it becomes. [The more, too, it trades away orientation to knowledge for access to information–TM.] The closer the professional is to the public, the more the service ideal survives and will continue to do so.

The Strengths and Weaknesses of Catalogs

In providing subject access, if there is one thing that library catalogs are good for it is in providing *overviews of search options* through displays of precoordinated subject headings and subject-subdivision strings. (Of course catalogs do other things, too.)

If there is one thing that they are notoriously bad for, it is in separating the wheat from the chaff–of pointing out the *best* individual sources from the many arrays and categories of options. (The fact that they point to professionally selected collections, however, puts them in marked contrast to Web search engines.) Library catalogs are also incompetent to lead readers to the best databases for journal articles among the hundreds available, let alone to the best articles themselves[14]; or to starting-point/orientation articles in the thousands of specialized encyclopedias that are not available online. Catalogs also have weaknesses in bringing to readers' attention government documents, microform research collections, and special collections. There are *other ways* to get into such things, however, as any good reference librarian knows. We don't need library catalogs to take on all of these functions–to "seamlessly" integrate "vast resources" all in one overwhelming source. The catalog is *one* necessary avenue of access to *some* necessary records; to overburden it with too many functions would be to kill a goose that lays golden eggs, and to undercut its ability to turn up *books* in particular. (Better home pages or portals that lead to the catalog *in relation to* other sources, could help here; but the catalog *itself* cannot lead seamlessly to *all* necessary sources–nor, for that matter, can even the best home pages or portals.)

The importance of seams

The larger point here is that visible "seams" among resources are in fact necessary for researchers. When a portal screen tells a researcher, in effect, to click *here* for access to books, *here* for journal articles, *here* for dissertations, *here* for Web sites, and *here* for newspaper articles, and so on–when it *shows the seams*, in other words, it thereby provides a structured overview of options that would otherwise be imperceptible. One of the greatest frustrations researchers have is that of not knowing "where they are"–of not knowing the extent of the results they initially retrieve, and whether they are looking at "everything." Seams between and among research options help readers to recognize a variety of paths that they can follow if their initial results are inadequate. Seams serve as perceptible *boundaries that provide points of reference*; without such boundaries, readers get "lost at sea" and don't know where they are in relation to anything else: they can't perceive either the extent of what they have, or of what they don't have.

Automated Collocation?

No other source–not *Books in Print* (with its inadequately subdivided subject headings), not Amazon.com, not Google–is as good at finding *books* by subject as a good library catalog. Automated means of subject collocation are no substitute for good cataloging. In Amazon.com, for example, the record for my own book, *The Oxford Guide to Library Research*, adds the following helpful notice:

> Customers who bought titles by Thomas Mann also bought titles by these authors:

- Franz Kafka
- J. K. Rowling
- Herman Hesse
- Andre Gide
- Feodor Dostoevsky

Much as I would wish to offer this as an example of the extraordinary insight, accuracy, and trustworthiness of the collocation software, I fear that more objective observers may reasonably conclude that a Large Mistake Has Been Made.

Catalogers Reading from Different Page?

(This is just an impression, but perhaps it's relevant: Much of the library world is trying to find reasons to induce people to continue coming inside the library's walls–*and pay their tax monies for supporting those walls and the nondigitized collections within them*–instead of just searching the Internet from their homes, schools, or offices. The cataloging wing of our profession, however, sometimes seems determined to create a product that will seamlessly cover "everything"–especially the Internet, which does not require entry within library walls–and do it in such a way that the catalog product itself can be tapped into by anyone, from anywhere, at any time. [The title of a recent conference of the New England Technical Services Librarians was "User Oriented Technical Services: All Things to All People."] It would help if catalogers would start thinking *outside the box of the Internet alone*, and realize how many important things–especially copyrighted books–are not *in* that Internet box, but still need good localized access and arrangement mechanisms. In other words, it might help to preserve libraries-as-places if catalogers were reading from the same page as the rest of us.)

Significant Differences Between OPAC Cataloging and Web Metadata: LCSH's Inextricable Links to LCC

Yet another function of cataloging–unlike metadata–is, again:

- **to serve as the functional index to the Library of Congress Classification scheme (LCC) in the classified bookstacks.**

It is through the subject headings in a library catalog, and their links to records with different class numbers, that researchers are enabled most efficiently to discover which areas of the stacks they need to go to (and which to avoid) for in-depth browsing of full texts of books on particular subjects. Without this linkage, which appears within catalogs themselves more than in the LCSH list (although the linkage is there, too, to a lesser extent), the exploitation of classified bookstacks would be greatly undercut, as it would not be easily determinable which stack areas cover which subjects. (Readers use library catalogs to index the bookstacks–there is no way they are going to endure catalogers'

indexes to LCC.)

The Continuing Need for Subject-Classified Bookstacks

The advantages of classified bookstacks are that they allow in-depth subject searching of *full-texts, not just catalog records*–i.e, readers can quickly scan whole groups of related texts *right next to each other*, not just for tables of contents, but also for running heads, illustrations, maps, charts, portraits, diagrams, statistical tables, highlighted boxes, typographical and color variations for emphasis, marginalia, footnotes, bibliographies, and indexes at the backs of books–none of which is digitized on catalog records. (Nor are the vast majority of the hundred thousand+ copyrighted books published each year making the "shift to digital" forms that Mr. Dillon apparently assumes; significantly, Mr. Dillon's own book itself has not made the shift.[15])

LCSH Must Function in Both Book and Web Contexts

The future of LCSH, in other words, must be planned with the maintenance of *this* context in mind, just as much as a Web context. Research libraries–unlike many special libraries–must continue to operate in both the contexts of online resources and print collections. It is not a matter of one context *rather than* the other, or one *superseding* the other, or one *shifting* to the other, or one *evolving into* the other. The requirements of discovering the knowledge contents of large book collections are not the same as those of searching the Web for unintegrated and unrelated information (which is, and probably will continue to be, the Web's primary–not only, but primary–function).

There are thus two contexts for the future use of LCSH, and the *book-collection context will not go away*. Nor can it be forced onto a Procrustean bed of postcoordinate search mechanisms more appropriate to the Web context without decimating the efficiency and "heavy lift" capacity of catalogs in providing subject access to large book collections.

This is, then, a real problem with some of the papers on the Bicentennial Conference Web site: They look at the future of LCSH *exclusively in the one context of Web resources*. (Pardon my redundancy; the point needs emphasis.) The "little error" of such a blinkered initial assumption will lead to "very serious consequences" for historians, biographers, literary scholars, and researchers in general who will, and often must, continue to use the vast stores of knowledge records, both retrospective and current, that simply are not and never will be digitized on the Web.

Missing Stakeholders

By the way, where are the representatives of stakeholders such as the American Historical Association, or the Organization of American Historians, the American Association of University Professors, or the associations of the other scholarly interests? If, by chance, the result of our Conference is to radically change the way *books* are given subject cataloging–so that future headings

no longer show up in browse displays related to existing headings; or so that the library catalog no longer functions as an index to the classified bookstacks–then shouldn't groups of professional academics who *depend on the book collections of research libraries* have a seat at the table? Surely we are not going to unilaterally declare that they will no longer need efficient subject access to large book collections in the future! How would *The Chronicle of Higher Education*, *Lingua Franca* and Nicholson Baker report such chutzpah?

Summary of Differences

It is highly unlikely that anyone will ever consider *library catalogs* as their first choice of entry into the Web–not at least, until library catalogs cover the *billion+* records indexed by Google et al. There are about 95,000 records in the RLG Union Catalog that point to digital resources (that is, having 856 fields)[16]; and we all hope this Conference will find ways to expedite the inclusion of still more such resources into library catalogs. But if we disregard, in our initial assumptions, the very features that make library catalogs such useful guides to substantive knowledge records then we will have done more damage than good to higher education in this country. Library catalogs and LCSH, unlike Web search engines with faceted metadata, have these features:

- They are tied, to begin with, to substantive, professionally selected records–books–that are proven media for conveying knowledge, not just information, and that can be economically preserved for centuries;
- They relate and link different subjects to each other in cross-disciplinary ways;
- They spell out the many unforseen aspects that lie (otherwise indistinguishably and unnoticeably) within any one subject field;
- They allow researchers to recognize relevant topics and relationships that they could never specify in advance;
- They guide researchers most efficiently to one or more areas of the bookstacks (rather than others), where so many of the substantive and non-digital knowledge records reside for quick browsing down to the page and paragraph levels.

The latter four functions are highly dependent on precoordination.

Blurred Distinctions

Two very important distinctions seem to be getting blurred in some of the papers before this conference:

1) Should the future of LCSH be considered primarily in terms of Web-type search softwares that do not allow precoordinate displays of subject strings–i.e., should it be our goal to change OPAC softwares themselves to be more like Google?

2) When we talk about extending LCSH to cover Web resources, do we mean:

(a) "covering" Web resources by creating surrogate catalog records for them, just as we do for books, which will show up "in the catalog"–i.e., within OPAC browse displays of precoordinated strings (as in the **Women–Services for** example at the beginning of this paper) in relation to the other surrogates already in the catalog?

Or do we mean:

(b) somehow adding LCSH elements directly to the headers of the actual Web records ("applications of metadata") out in the Internet–i.e., to headers residing within the Web sites themselves, not to surrogates merely pointing to them from their residence in the OPAC?

Intellectual Property Issues

Regarding (1): Given the billion+ Web sites that already exist, and the Web's rate of growth, isn't it just common sense to regard the application by catalogers of LCSH metadata elements to the headers of Web records, directly, to be a hopelessly Sisyphean task? Isn't it common sense also, to begin with, to recognize that *we do not have the authority to tamper directly with the intellectual property of billions of Webmasters by obtruding our presence into their sites*? We can do anything we want with *surrogate catalog records that we create in our own OPACs*; but we simply have no right to tamper directly with the metadata on headers within Web records themselves.

Perhaps, then, we can suggest improvements, not to the countless Webmasters' sites themselves, but to the commercial engines like Google and NorthernLight, et al., which *index* those sites. That is, perhaps we can recommend ways in which their weighting and ranking softwares can be tied to authority lists, in order to map words in retrieval results to faceted LCSH elements, which would provide some additional measure of control to the keyword-weighting process. (Precoordinated strings would be out of the question in this context–no machine could assign them automatically.)

I have no objection whatever to our making suggestions to the search engines that we do not control ourselves. But in the blur of these distinctions, I would emphatically remind everyone, again, that intellectual property rights are involved: *librarians do not and cannot control these*

commercial Google-type indexing enterprises any more than we can control the Webmaster-created headers of the Web sites they index.

Merging OPAC Searching with Internet Searching?

The only things we can *control* are the things we create ourselves. This means *library catalogs*, not Google or HotBot or their commercial cousins. If we confine ourselves to examining the future of library catalogs–the only things we can control–then we have different options:

Option A: We can attempt to merge the searching of library OPACs with the searching of Internet sites through software changes. This merging could theoretically be done "from the outside in," or "from the inside out":

A.1. "From the outside in." We could abandon our existing OPAC softwares for searching bodies of catalog records separated from the Web. By merging our catalogs into the Web we could open their full contents directly to Web search engines such as Google or Yahoo. We could simply piggyback on these existing services already known to, and widely used by, researchers. A Google search of the future, then, would seamlessly turn up surrogate catalog records for books, created by librarians, in the same operations that retrieve Web sites created by others. We could continue to assign LCSH elements that would serve as metadata elements searchable by Google type engines rather than by segregated OPAC softwares. Since Web engines cannot show precoordinated strings in browse displays, we should simply abandon precoordination in LCSH.

A.2. " From the inside out." We could radically change our own library catalogs so that they, *like* Google, try automatically to index not just local collections-within-walls but the entire Web, via spiders, crawlers, harvesters, and term-weighters *of our own creation*. Unlike Google, however, *our* automated indexes could add faceted LCSH elements through softwares that would map weighted keywords to controlled LCSH elements, whether or not these elements appear in the headers or bodies of the indexed sites Web sites themselves that exist beyond our own catalog records. While, for intellectual property reasons, we could not force LCSH elements into the headers of Web sites created by others, our software could add them to the displayed *results* of weighted keyword searches, to provide additional elements of control not otherwise present. This option, too, would necessarily abandon the display of precoordinated strings of LCSH terms, because no mapping software could possibly create proper strings, or displayed linkages among them, simply on the basis of weighted keywords.

If we go in the direction of **Option A**, in either of its variants, we would effectively have to merge catalog records for books–which we would continue to create–into the same "pool" as the Web environment that we seek to catalog, and which already exists outside

our present catalogs. The major difference lies in whether we search the records by existing external softwares (from the outside in) or through new softwares of our own devising (from the inside out). In neither case would there be any point to continuing precoordination in LCSH, since neither option would be capable of showing subject heading strings in browse displays.

Book Records Buried in Chaff, Loss of Connection Between LCSH and LCC

Before considering option (B) for the future of library catalogs, let me say why I think option (A) is unworkable. First and foremost, even if faceted LCSH terms were somehow mapped automatically to all Web sites *and* added manually by catalogers to individual book sites, the book records would become so buried within the overwhelming chaff of the Web that researchers would no longer be able even to identify the ones most relevant to their topics. Nor would researchers be able to view such records for books in relationship to other book records–or, for that matter, identify books in relation to the most relevant Web sites.

There would just be too much chaff; and the assignment of faceted LCSH elements would simply not be enough to control retrieval in any way noticeably better than what Google does.

Such a Web-search library catalog would utterly sever the existing network of strong connections *from* book records cataloged with precoordinated LCSH elements *to* particular areas of their local classified book collections. This would effectively vitiate the possibility of scholars efficiently browsing classified book collections locally.

I think we may reasonably conclude that future catalogs should not, like Google or Hotbot, try to swallow the whole Internet or to merge into it; they will maintain their utility only by indexing highly-selected portions of the Web, and in a way that does not overwhelm researchers with unwanted chaff.

Expanding the Range of Free-Floating Form Subdivisions to Include Web Sites

A second option for the future of library catalogs would be:

Option B: We could continue to use the software of existing library catalogs that show browse displays of precoordinated LCSH headings, but expand the range of (free-floating) subdivisions to include form subdivisions for Web sites. Let me repeat here the example given earlier:

Women–Services for
Women–Services for–Bolivia–Directories
Women–Services for–Caribbean area–Case studies
Women–Service for–Ethiopia–Congresses

Women–Services for–Germany–History
Women–Services for–Michigan–Evaluation
Women–Services for–New Zealand–Bibliography
Women–Services for–North Carolina–Finance
Women–Services for–Study and teaching–United States
Women–Services for–Study and teaching–United States–Web sites (.edu)
Women–Services for–United States–Directories
Women–Services for–United States–Web sites (.com)
Women–Services for–United States–Web sites (.edu)
Women–Services for–United States–Web sites (.edu)–Data archives
[Again, "Data archives" may not be an appropriate subdivision for this particular subject; I offer it here just as a pattern example.]
Women–Services for–United States–Web sites (.edu)–Discussion lists
Women–Services for–United States–Web sites (.edu)–Portals
[Again, "–Site directories" might be an alternative, in which case a cross-reference is needed: Site directories USE Portals]
Women–Services for–United States–Web sites (.gov)
Women–Services for–United States–Web sites (.org)
Women–Services for–Wisconsin–Periodicals
Women–Services for–Zambia–Directories

Of course, live links would be provided from the catalog surrogates to the actual Web sites, insofar as licensing agreements allow.

Precoordinated displays like this in OPACs would (1) separate the substantive Web sites from the clutter of chaff turned up by Web search engines, and (2) show them in relationship to scholarly book records–an ideal outcome. We would be using precisely the strengths of the catalog in its unique display potential, as well as in its selectivity, to overcome the weaknesses of the Web. These goals ought to be at least part of what we are aiming for.

The Large Question

But we need to do more than just this. **The larger question before this Conference, I think, is this: How can we (a) simultaneously get LCSH into both metadata fields of Web records created by other people *and* into the OPACs that we create ourselves; and (b) do it in a way that will simultaneously exploit the strengths of both the flexible postcoordinating software of existing Web search engines *and* the powerful browse screen capabilities of OPACs?** This would be **Option C**, to which I shall return.

Is Loss of Precoordination Really Logical?

As a prelude to **Option C**, however, I must first comment directly on Lois Mai Chan's paper.[17] When Ms. Chan asks the question "What direction and steps need to be taken for LCSH to overcome these limitations and remain useful in its traditional roles as well as to accomodate other uses?" she specifically includes "systems with index browsing capability" among the "limitations" that must be "overcome. " She reports, further, on one of her current projects:

> Using LCSH as the source vocabulary, FAST (Faceted Application of Subject Terminology), a current OCLC research project, explores the possibility and feasibility of a postcoordinate approach by separating time, space, and form data from the subject heading string (Chan et al. in press).

She also comments, a paragraph later:

> Considering the gradual steps the Library of Congress has taken over the years, even a person not familiar with the history of LCSH must conclude logically that LCSH is heading in the direction of becoming a fully faceted vocabulary. It is not there yet; but, with further effort . . .

The phrase "not there yet" obviously implies an acceptance, and recommendation, of what seems to be a "logically" inevitable transformation of LCSH into a system of fully faceted elements (which can only be contrasted with a system of precoordinated strings). These comments, however, need to be placed in the context of another very recent paper by Ms. Chan, appearing in *Cataloging & Classification Quarterly,*[18] in which she writes:

> Within the OPAC environment, where trained personnel is available for the creation and maintenance of complex subject heading strings and the online system is capable of handling such, the current rules and policies for complex syntax can continue to function.

Amen. This point, I think, needs much greater emphasis than it receives in Ms. Chan's paper before the present Conference. The **Option C** that I will propose is one that I think (hope?) we can agree on; but here is the key point: we must consider the future of LCSH, as I have argued above, in *two* continuing environments that are *very different* from each other: one, the OPAC/book-collection environment, and the other, the Web/networked environment. And *because* the book collection environment will not transform, merge, or evolve into the Web/networked environment but will always remain distinct from it, I maintain that we need a future LCSH that does not lose the many existing strengths of precoordinate displays. This is the crucial difference: one environment supports the display of precoordinate LCSH strings and the other simply does not.

What I am afraid of is that Ms. Chan's conference paper readily lends itself to misinterpretation, because while it does indeed recognize (some) important distinctions between the two environments, its portrayal of the "logical" future of LCSH in the Web/networked environment

silently entails its loss of precoordination in the OPAC/book collection environment–unless Ms. Chan advocates that two different LCSH systems be maintained in the future for the two different environments. She is silent on this; but I suspect she (and everyone else) would regard the maintenance of two different LCSH systems to be economically as well as intellectually unsupportable.

What Would Be Lost

The theoretically extrapolated loss of precoordination, however, is neither logical, nor necessary, nor inevitable, nor desirable:

- It is not logical to abandon precoordination when the *very meaning* of so many LCSH terms is dependent on the word-order of their phrasing, in ways that cannot be recaptured by postcoordinate Boolean combinations *or* by word-proximity searches that drop out relational prepositions as stopwords.

- It is not logical to abandon precoordination when to do so would uproot tens of thousands of LCSH strings from a *vast web of specific linkages to LCC*–i.e., changes in the word order of the subject strings *also* changes the classification areas to which they point.

- It is not logical to abandon precoordination when browse displays of subject-string phrases *enable researchers simply to recognize* whole ranges of options that they could never specify in advance through postcoordinate combinations (e.g., **Yugoslavia–Antiquities** rather than just **Yugoslavia** AND **History**; **Afro-American whalers** rather than just **Afro-Americans** AND **History**; **Greek language–Onomatopoeic words** rather than just **Animal sounds**). The larger the file, the more researchers are dependent on *recognition* of options that they cannot articulate beforehand.

- It is not logical to abandon precoordination when the existence of the *vast cross-reference structure* between and among headings is so heavily dependent on the retention of ordered strings (e.g, **Women-Psychology** NT **Leadership in women**; **Afro-Americans–Education** NT **Segregation in higher education–United States**).

- It is not logical to abandon precoordination when the *relationships of alphabetically-adjacent headings* within the thesaurus would be entirely lost without it (e.g., **Monasteries** is linked to scores of precoordinated neighbor headings such as **Monasteries and state** and **Monastic and religious life of women** simply by their displayed contiguity rather than by any formal cross-references).

- It is not logical to abandon precoordination when LCSH, unlike any other thesaurus, must simultaneously cover *all subject areas–not just one, as other thesauri do–and show relationships among them* that readers could not specify in advance.

Nine years ago Ms. Chan read a paper to the Airlee House Conference, similarly calling for less precoordination and greater use of postcoordinate combinations of individual, faceted elements in LCSH. The members of that conference listened respectfully, but then ignored the substance of the paper–i.e., the course of the subsequent discussion immediately became, effectively, not "Should there be less precoordination?" but rather "*Given the need to retain precoordination* [for the above reasons], what should be *the order of the string elements*?" Subsequent improvements in search software–as in Google, Hotbot, et al., which did not exist at the time–have not invalidated any of the above reasons for retaining precoordination in LCSH.

A theoretically-extrapolated projection of greater postcoordination of individual facets simply ignores the reality of the many functions LCSH already serves in the real world of real library collections; and these continuing (and growing) functions are just as much a part of its history as is the trend to break phrase headings into subdivided (*but still precoordinated*) strings in browse displays. The real world of practice and function puts *real and definite limits* on the "direction" of LCSH toward "becoming a fully faceted vocabulary." None of these realities is given anything more than passing mention–most are not even mentioned at all–in Ms. Chan's current paper. This will never do; such "errors in the beginning lead to serious consequences in the end."

While greater facetization–if there is such a word–of LCSH may indeed be a desirable goal *in a Web environment such as* **Option A** *above*, in which we abandon our current OPAC softwares, I think we need to question whether **Option A** is even possible, let alone desirable, to begin with. One crucial point is that Ms. Chan simply does not consider the question of intellectual property: Can librarians add LCSH elements to headers of countless Web sites whose Webmasters have no obligation whatever to pay any attention to what librarians want? Answer: No, we cannot. If, then, librarians cannot obtrude our terms into other people's intellectual property sites, what chance do we have of getting independent Webmasters to voluntarily start using LCSH elements in their headers? And what will the results of LCSH, either faceted or precoordinated, applied by rank amateurs be like? Will it sustain the "heavy lifting" capacity that our large research libraries–and our nation's intellectual culture itself–require? The results of utterly fragmented LCSH elements applied as metadata to Web headers by amateurs, I suspect, would hardly bear any relation to what professionals usually think of as "vocabulary control." (And how do we prevent Webmasters of porno sites from having a field day with their voluntary use of LCSH's **Women** terms in their headers?)

The larger question here, of course, is this: Should our profession consider the primary future use of LCSH to be by Webmasters *over whom we have no control*? I think not.

Getting Librarian-Created LCSH Elements Into the Headers of Web Sites

But I also think there *is* a way that we can get professional-librarian-assigned LCSH elements into the headers maintained by independent Webmasters. This is a proposal is similar to what Regina

Reynolds recommends in her paper, and in line with what Priscilla Caplan recommends when she calls us to "work proactively with publishers."[19] It is:

> **Option C: an Online Cataloging In Publication (OCIP) program that mirrors our current CIP program for printed books.** With such librarian-created metadata added to the Web sites of *quality-screened participants*, we would have the best of both environments: We could continue to assign LCSH in traditional precoordinated strings on the *surrogate records that we create*–but these records would then appear *in both environments*: directly within the program's Web records as metadata in their headers *and* simultaneously in OPACs as catalog records.
>
> In the Web environment, as metadata, even if the LCSH elements are assigned as strings, their individual words or facets could still be searched postcoordinately by *existing* services such as Google and NorthernLight, without our having to overhaul our own expensive catalog softwares.
>
> In the OPAC environment, in contrast, the same LCSH elements could still be searched in their precoordinated forms in catalog browse displays–as well as postcoordinately. Their preocoodinated display, as with the **Women–Services for** example above, would *relate* the quality-selected Web sites to existing and future *book* records, as well as to other quality-selected Web sites–and also do it is a way that does not undercut the widespread linkages of LCSH to LCC in the classified bookstacks, nor undercut the cross-reference structure, undercut users' recognition capabilities, etc., etc.
>
> I do not mean to suggest that library catalogers should create catalog records *only* for Webmasters who sign up for the OCIP program. Far from it. Library catalogers should be free to create surrogate catalog records in their OPACs that point to any Web site at all worthy of being brought to researchers' attention. And there is nothing in this OCIP proposal to prevent this. The *extra* advantage of an OCIP program, however, is that the cataloging data created for participants in the program would *also* become searchable as metadata in the participants Web sites–i.e., accessible not just on catalog surrogates through library OPACs but also within metadata fields accessible via Google and HotBot and all the other engines.
>
> This proposal also has the advantage of saving us the expense of radically redesigning the expensive search softwares of our existing OPACs. And it includes all of the strengths of **Option B** while also averting the intellectual property problems, and those of overwhelming chaff, in **Option A**.
>
> Yet another likely advantage: if the existence of the OCIP program were made known as widely among Webmasters–especially corporate bodies–as CIP is among

publishers, then the Webmasters of high-quality sites will probably start trying to bring their sites to our attention, on their own initiative. Just as CIP records make books more attractive to libraries, OCIP records would make Web sites similarly attractive. To get into the program, however, Webmasters would have to document both the quality and the likely longevity of their sites for us. That means librarians wouldn't have to spend endless time surfing around, *looking* for the best sites. Their producers would strive to bring them to our attention.

Of course there is a larger managerial/administrative problem to be worked out: Should the Library of Congress be the *only* library responsible for creating OCIP records, as with CIP records? I think this is inadvisable. Given the sheer size of the Web, and the number of possible applicants for participation in the program, the work would have to be divvied up. I think that can be managed. (Perhaps division by States, with first priority given within them to local .edu domain sites? [LC could concentrate on federal .gov sites.] A State-run OCIP program, administered through both State libraries and State University libraries, might also enable us to get a handle on how to divvy up electronic *preservation* responsibilities. We can't even begin to preserve everything on the Web; but perhaps the sites of OCIP participants within each State would provide an initial rough focus for preservation attention? Indeed, an increased likelihood of preservation might well serve as an incentive for Webmasters to join the program.) The details are outside the scope of this paper, and probably outside my own competence to imagine.

Doesn't **Option C**, however, address many of the major problems confronting this Conference? Priscilla Kaplan says in her paper, "The most critical factor in the future of DCMES [Dublin Core Metadata Element Set] is whether a working organization can be achieved to manage the change process and to produce the documentation, support structures, and policies required by an international community of implementers holding very little in common."[20] I suspect an OCIP program–probably having to extend beyond U.S. States to foreign participants–holds the best hope of creating the *locus* sites that will be necessary to create these support structures.

The Need for Consistency and Accuracy in Subject Heading Assignment

There is one further issue that I think this Conference needs to address squarely: If we are going to use LCSH in *both* OPAC and Web environments of the future–and I heartily hope that we will–it really does make a difference that we strive for consistency and accuracy of subject-heading assignment. There isn't any "control" in "vocabulary control" to begin with if subject cataloging is relegated to low level technicians who know nothing of specific entry or cross-references. Nor can there be much control if we regard Web sites rather than books as the primary targets of our cataloging activities, for the simple reason that LCSH elements appearing in metadata fields, *if considered only as separate from OPAC displays of the same data*, do not require the many extra controls of precoordination, cross-referencing, links to LCC, or displayed alphabetical adjacency to related headings.

Are Web Sites More Important Than Books?

This, then, brings us to some of the proposals put forward by my former LC colleague Sarah Thomas, which she makes in her paper, "The Catalog as Portal to the Internet."[21] There are many, many worthwhile observations in this paper. But then it comes to:

> 1. We should decisively reduce the amount of time we devote to the cataloging of books in order to reallocate the time of our bibliographic control experts to provide access to other resources, especially Internet resources . . .

I thank Ms. Thomas for a bluntness comparable to Mr. Dillon's. It is easier to engage in healthy debate when one's assumptions are not buried as concealed propositions. The forthright message here is that books are now of less importance to our culture than are Internet sites.

I beg to differ.

In the first place, our larger culture depends on libraries and librarians to provide *free* access to *books*. The full-texts of most books are not on the Internet, and most never will be, for copyright (life of author plus seventy years) and preservation reasons alone. Those that do appear, either freely available to anyone from anywhere, or free only to users of site-licensed terminals within library walls, will not be *read* online because of their lengths, but will be printed out individually at much greater-than-present costs either to libraries or to the environment, or both.

Further, it will very soon be the case that no one–not even poor people–will be dependent on libraries or librarians for access to the freely-accessible portions of the Internet[22]; but our culture as a whole will still be very much dependent on libraries and librarians for *free* access to the scores of thousands of *books* that continue to be published every year (cf. *Bowker Annual*), as well as to the low-use texts of earlier decades and centuries.

Further, all of those home- and office-connected Internet searchers will not be dependent in any way on libraries or library catalogs for ways to search the Internet: they will have Google, Hotbot, AltaVista, NorthernLight, and a wide array of other avenues of access freely available to them. Even if librarian-created catalogs are modified to include selected high-quality Internet sites (as in Options B and C above), I think it is highly unlikely that searchers would consider them as their first or most important avenues of access to the Net, in preference to Google et al. The virtue of library catalogs will lie precisely in:

> **(a)** pointing researchers to important resources–*books*–that cannot be found on the Net to begin with;
> **(b)** pointing them to high-quality Net sites that will otherwise be buried in the chaff turned up by Web search engines; and

(c) in relating books and quality Web sites to each other intelligibly rather than haphazardly.

But researchers will lose out on the benefits of **(a)** and **(c)** exactly to the extent that librarians, following Ms. Thomas's advice, "decisively ... reallocate" their time and attention to **(b)**. It seems that Ms. Thomas does not consider **(a)** and **(c)** as important to begin with. As a reference librarian who must help thousands of very confused researchers every year, I beg to differ. I do consider them very important.

The additional danger of slanting library catalogs primarily to Internet sites has already been alluded to (pp. 11-12 and footnote 7): We librarians and information specialists may unintentionally wind up dumbing down our larger culture if we give the primacy of our attention to a resource that is itself slanted to shorter (rather than longer) texts, visual images, audio resources, and graphical displays over textual explanations–i.e., to a medium that much more readily conveys data and information than knowledge or understanding. Again, our larger culture does not depend on librarians or library catalogs for free access to the Internet; but it very much does depend on us for free access to the substantive *alternatives* to the Net, and for the integration of the Net *into larger webs of knowledge relationships*. These needs cannot be met under Ms. Thomas's proposal for redefining our priorities.

Accepting Copy Cataloging "with little or no modification"?

Ms. Thomas then goes on to say:

> 2. In order to reduce the time spent cataloging books, we will need to investigate and implement a combination of the following:
>
> * * *
>
> Accepting copy cataloging with little or no modification from other cataloging agencies, including vendors

Ms. Thomas's enthusiasm for accepting virtually any copy cataloging "with little or no modification" has a noteworthy history. It was she who led the Library of Congress into adopting this practice in a big way. (Even now, however, it is not easy to generalize about LC's cataloging operations; there are about three dozen cataloging teams, and they vary in the level of review that they give to copied records. Some do accept copy "with little or no modification"; some don't.)

"Only about 20% agreement among catalogers"?

Ms. Thomas, in order to embark LC on the project of accepting copy-cataloging widely, invited her friend and colleague Carol Mandel, from Columbia, to address LC's troops in a Cataloging Forum meeting on 12/9/1993. There Ms. Mandel told all of us, with Ms. Thomas's approval, that

"studies" show that there is "only about 20% agreement among catalogers" concerning which subject headings should be assigned. This assertion repeated Ms. Mandel's claim in her 1991 "Cataloging Must Change!" article in *Library Journal*,[23] written with Dorothy Gregor. Because of this alleged "lack of interindexer consistency," this articles says, "Catalogers can be more accepting of variations in subject choices in member copy and need not spend undue time determining whether their analyses are consistent with LC's and with those of catalogers elsewhere." Evidently on the basis of Ms. Mandel's scholarship and sources cited , Ms. Thomas herself wrote in 1993: "Recent studies have determined that intersearcher consistency does not exist. . . . *With this new knowledge*, administrators and catalogers are asking to what extent strict consistency of application of subject headings increases the quality of the bibliographic record for use by end users"[24] [emphasis added].

The claim that there is only 20% agreement among subject catalogers was simply accepted as "knowledge" by Ms. Thomas. LC's acceptance of cataloging copy–with subject headings largely unchecked for accuracy, completeness, or consistency–shot up from 1,800 titles in 1991 to over 45,000 in 1994, under her direction.

A few years later, having come across a number of disturbingly inaccurate records that I found too late to help a few readers who could have profited from them, I began to wonder about the basis of Ms. Thomas's faith in copy cataloging that is accepted with little or no modification. So I went back to Ms. Mandel's "Cataloging Must Change!" article to check out its footnotes.

Getting the Basic Facts Wrong

What I found, briefly, is that Ms. Mandel and co-author Ms. Gregor had their facts 180 degrees backward: the studies they rely on show that the low interindexer consistency rate of ca. 20% shows up repeatedly precisely *in the absence of* vocabulary control mechanisms.[25] This is the figure achieved by amateurs who are trying to guess which keywords should be used to index a document, usually in situations entirely lacking thesauri, cross-references, familiarity with cataloging principles (especially the convention of specific entry), and established catalogs exhibiting an established pool of vocabulary-controlled records. Subsequent studies suggest that ca. 80% consistency can be expected among professional catalogers who follow the rules.[26] One, by Elaine Svenonius and Dorothy McGarry, states: "*The price that is currently being paid for lack of subject expertise in non-LC subject cataloging is that over 50 percent of the books so cataloged [i.e., by agencies other than LC] are either missing headings or have headings that are incorrect, dated, or questionable*"[27] [emphasis added].

Result of "little or no modification" in Subject Cataloging: *Subject Guide to Books in Print* Example

What is the result, for users, of bad subject cataloging? Since Ms. Thomas herself appeals to anecdotal evidence in her own paper, I will have no qualms in using it here. I would appeal to it in any event; the importance of examples, case studies, and first-hand testimony is established in many fields, including Law, beyond our own discipline.

Let's look first at subject cataloging from a commercial source. One that is readily available in libraries throughout the country is Bowker's *Subject Guide to Books in Print* (*SGBIP*). To stay within the ballpark of the **Women** examples used elsewhere in this paper, here are five examples of the subject cataloging done by the Library of Congress and *SGBIP*:

- Title: *The Beijing Declaration and the Platform for Action: Fourth World Conference on Women, Beijing, China, 4-15 September 1995.*

  LC headings: World Conference on Women (4th : 1995 : Beijing, China)
  Women–Social conditions–Congresses
  Women's rights–International cooperation–Congresses
  Women in development–International cooperation–Congresses

  SGBIP: Women

- Title: *Women as Elders: The Feminist Politics of Aging*

  LC Headings: Aged women–Congresses
  Aged women–Religious life–Congresses

  SGBIP: Women

- Title: *Female Gangs in America: Essays on Girls, Gangs and Gender*

  LC Headings: Gangs–United States
  Female juvenile delinquents–United States
  Female offenders–United States

  SGBIP: Gangs

- Title: *The Women, Gender and Development Reader*

  LC Headings: Women in development
  Women–Social conditions
  Women–Economic conditions
  Women–Developing countries

SGBIP: Women

- Title: Women Overseas: Memoirs of the Canadian Red Cross Group

LC Headings: Canadian Red Cross Society–Biography
World War, 1939-1945–War work–Red Cross
Korean War, 1950-1953–Participation, female
World War, 1939-1945–Personal narratives, Canadian
World War, 1939-1945–Participation, female
Korean War, 1950-1953–Personal narratives, Canadian
Nurses–Canada–Biography

SGBIP: Red Cross
Women
Canada

Should commercially-available subject cataloging such as this from *Subject Guide to Books in Print* be accepted "with little or no modification"? Subject cataloging like this provides virtually no "control" at all, and virtually no possibility of readers' recognizing such headings within meaningful relationships. Note that the LC subject-strings would all show up intelligibly within larger browse screens, displaying other subdivision-aspects of the same topics in immediate proximity.

Result of "little or no modification" in Subject Cataloging: Unreviewed Cataloging from Bibliographic Utilities

What about the non-LC subject cataloging available from bibliographic utilities–the kind that Svenonius and McGarry found to be inaccurate or incomplete half the time? Again, the evidence is anecdotal; most reference librarians and catalogers just don't have the time to do statistical studies like Svenonius/McGarry.

Cataloger Jan Herd gave me an example she described as "not unusual in the books I receive." The title of the work was *The Credit Repair Rip-Off: How to Avoid the Scams and Do It Yourself.* The subject headings supplied by the copy cataloging were:

1. Debtor and creditor–United States
2. Debt relief–United States

Ms. Herd wrote to me:

> The first heading is a "law heading" and classes in KF1501 according to a law cataloger here in [this division]. [Note this cataloger's immediate recognition of the need for a proper tie to be established between LCSH and LCC.] He stated it should not be used on this book since it is not in scope as a law book. The second heading is also not appropriate for this book since Debt relief refers to macroeconomics . . . country level debt relief, renegotiation, etc.
>
> I received the book . . . . I had to change the headings to:
>
> 1. Consumer credit–United States
> 2. Credit ratings–United States
>
> The book was classed in HG3756 which corresponds to Consumer credit by country.
>
> This type of wrong thinking in assigning subject headings is not unusual in the books I receive. . . . When we multiply this kind of work on a daily basis we are polluting our database rapidly. We need a library EPA to impose "environmental impact charges" on libraries contributing to the pollution.

Usually I don't write down examples of bad copy cataloging unless there's a compelling reason; I have many other things to be doing with my time, and I generally just have to rely on what catalogers provide. Often, too, by the time I discover that I've overlooked some good sources due to their not showing up under the right headings, the reader who needs the books has vanished. I did write down an example, however, that was brought to my attention two months ago. A colleague of mine who is a rare book and manuscript cataloger in a private collection found, to her dismay, that her own scholarship was undercut by inadequate copy cataloging accepted by LC.

Result of "little or no modification" in Subject Cataloging: Undercutting Overviews Needed by Scholars

Dr. Melissa Conway's book, *The Diario of the Printing Press of San Jacopo di Ripoli, 1476-1484: Commentary and Transcription* (Firenze: L. S. Olschki, 1999), was published last year; and recently she was given an advance copy of a review of the book that will appear in 2001 in the journal *Book Collector*. Most of the review is irrelevant here, but on one point its writer faulted Dr. Conway's historical survey for not being updated by a particular book in the field that, the reviewer says, she should have read. Conway had been monitoring the appearance of books in the relevant field by regularly checking LC's catalog for works under the headings that had been applied to a standard work that she did make use of, Christian Bec's *Les Livres des Florins (1413-1608)*. The subject headings assigned to this book are:

> **Books and reading–Italy–Florence–History**
> **Libraries–Italy–Florence–History–1400-1600**
> **Libraries–Italy–Florence–Catalogs**

Florence (Italy)–Intellectual life

The book she is criticized for overlooking is Armando F. Verde's *Libra tra le Paret Domestiche*; this work itself is a kind of supplement to an earlier work by Verde, *Lo Studio Fiorentino, 1473-1503*. Evidently the non-LC cataloger who created the record for the *Libri* book didn't look at its contents carefully, but simply assigned to it the one subject heading given previously to the *Studio* record:

Universita di Firenze–History

In other words, according to Dr. Conway (who is herself a professional cataloger), the *Libri* book does indeed cover the subjects of **Books and reading** and **Libraries** in **Florence**, but the subject headings that ought to have indicated this were never assigned by the non-LC cataloger. And LC accepted the one inadequate subject heading "as is."

The ultimate point is that a serious scholar relied on a subject search of LC's catalog to do the "heavy lifting" it is supposed to do: not just to give her "something" on her topic, but rather to provide an *overview of the range* of significant, relevant resources available. And inadequate copycat subject cataloging, accepted with no modification, undercut that goal.

I do not mean to suggest that Dr. Conway's career is threatened as a result of inadequate subject cataloging; on the other hand, she is not in an academic position requiring "publish or perish" output, to begin with, or favorable reviews of it. An academic whose tenure is on the line in a similar situation, however, may have much stronger feelings about a library catalog that is supposed to, but doesn't, do the "heavy lifting" that a serious scholar expects of it.

The Need for Quality Subject Cataloging

And so I must beg to differ with Ms. Thomas's rather abrupt dismissal of the value of quality cataloging, which simply cannot be taken "with little or no modification" from the existing pools of ever-decreasing professional work.[28] Copy cataloging of subject headings and class numbers–if it is truly going to help library catalogs accomplish what scholars *need* to have accomplished–does indeed have to be checked with an eye to consistency, completeness, relationship, and accuracy. I realize, of course, that if Ms. Thomas is still promoting an opposite view in the wake of the Svenonius/McGarry study, and in the wake of the exposure of the factually false premises of the Mandel/Gregor article that she unquestioningly accepted as "knowledge," then nothing added here is likely to change her mind. But I sincerely hope that other participants in this Conference will realize that good subject cataloging–precoordinated, browse-displayed, linked to LCC, cross-referenced, and at specific levels–does indeed make all the difference in the world when its goal is understood to be that of providing *structured overviews of the range of significant sources relevant to a topic, rather*

than just "something"–i.e., rather than just isolated and unintegrated information.

I'll say it again: If we as professionals are not making *knowledge* more available than it would be without our efforts–knowledge in its largest possible frameworks of relationships, interconnections, and linkages–rather than just isolated bits of *information*, then we are not fulfilling the most important responsibilities we have to our larger culture.

1. Mortimer Adler, *Ten Philosophical Mistakes* (New York: Macmillan. 1985), xiii.

2. LC itself has closed stacks, at least under its current administration; but most libraries using LCSH and LCC have open stacks in which this information would be immediately useful.

3. Note that Lois Mai Chan's Faceted Application of Subject Terminology (FAST), discussed in her "Expoiting LCSH" paper at <http://lcweb.loc.gov/catdir/bibcontrol/chan.html>, would, if applied to LCSH in both Web and OPAC environments, simply destroy the linkage of such strings to definite LCC stack areas. The same LCSH system, in other words, could not be used in both environments without great damage being done in the OPAC context, because postcoordination of the geographical "space" elements would destroy the indexing significance of the ordered string's link to LCC.

4. Unfortunately, the need for maintaining subject-classified bookstacks themselves seems to have dropped off the radar screens of many writers in our field. The continuing need for such classified shelving, and the reasons that it cannot be replaced by searching by class numbers within computer catalogs, are discussed at length in my paper, "Height Shelving Threat to the Nation's Libraries" at <http://studentorg.cua.edu/slislab/shelving.htm>. It also contains a discussion of the false notion that an "evolution" to digital forms is "inevitable." (In subsequent developments at LC, the matter seems to have gone into hibernation; the threat is no longer immediate.)

5. Numerous other examples can be found in the same book, as well as in the subsequent *Oxford Guide to Library Research* (Oxford U. Press, 1998).

6. Again, the FAST agenda (cf. note 3 above) would destroy such networks of cross-references if a scheme usable for LCSH in the Web environment were simultaneously forced onto LCSH in the OPAC environment. Since two separate LCSH systems cannot be reasonably maintained, the value of any proposed improvement needs to be critically examined for its impact in *both* environments. One hopes Ms. Chan's forthcoming study will address rather than ignore this crucial issue.

7. The evidence is not strong enough to establish a direct cause-and-effect relationship, but the observations made in a recent *Washington Post* article (April 26, 2000) by reporter Linton Weeks are not such that librarians and information professionals can simply ignore warning signs that are all around us, such as: "In the August 1999 issue of Conservation Biology, David W. Orr, a professor at Oberlin College, wrote that the human vocabulary is shrinking. By one reckoning, he observed, the working vocabulary of 14-year-olds in America has plummeted from 25,000 words in 1950 to 10,000 words

today. 'There has been a precipitous decline in language facility,' says Orr. 'This is nothing less than a cultural disaster.'" Weeks also quotes Keith Devlin, identified as dean of science at St. Mary's College in California and a senior researcher at Stanford; according to Devlin, "We may be moving toward a generation that is cognitively unable to acquire information efficiently by reading a paragraph. They can read words or sentences–such as bits of text you find on a graphical display on a Web page–but they are not equipped to assimilate structured information that requires a paragraph to get across. . . . Half a century after the dawn of the television age, and a decade into the Internet, it's perhaps not surprising that the medium for acquiring information [that a large number of the 10,000 college students surveyed] find most natural is visual nonverbal: pictures, videos, illustrations and diagrams." The dumbing down of learning–the loss of larger knowledge frameworks in our culture–is also commented on by Vladimir N. Garkov, "Cultural *Or* Scientific Literacy?," *Academic Questions*, 13, 3 (Summer, 2000), pp. 63-64: "A report on the first national assessment of our 17-year-old students' knowledge of history and literature found that this 'nationally represented sample of eleventh-grade students earns failing marks in both subjects.' A more recent study on cultural literacy, reported in the *Chronicle of Higher Education* (14 June 1996) found that only 7 percent of our graduating college students answered fifteen or more of the twenty questions correctly. The results from the National Assessment of Educational progress history exam show that only four out of ten high-school seniors demonstrated even a rudimentary knowledge of their own American history." Garkov cites Diane Ravitch and Chester E. Finn, Jr., "What Do Our 17-Year-Olds Know? A Report on the First National Assessment of History and Literature (New York: Harper & Row, 1987); Study on cultural literacy, *Chronicle of Higher Education*, 14 June, 1996; and L. Hancock and P. Wingert, "A Mixed Report Card," *Newsweek*, 13 November, 1995, 69.

8. Walt Crawford and Michael Gorman, *Future Libraries: Dreams, Madness, and Reality* (Chicago: American Library Association, 1995), p. 5; emphasis in original.

9. Martin Dillon, "Metadata for Web Resources: How Metadata Works on the Web." <http://lcweb.loc.gov/catdir/bibcontrol/dillon_paper.html>

10. F. W. Lancaster, "Second Thoughts on the Paperless Society," *Library Journal,* 124, 15 (September 15, 1999), 48-50.

11. Walt Crawford, "Paper Persists: Why Physical Library Collections Still Matter," *Online*, 22, 1 (1998), 42-48.

12. Dillon, "Metadata" (ibid.).

13. Lancaster, *ibid.*

14. State-of-the-art or overview "review" articles are especially prized by researchers. But it takes reference librarians to point out both the very existence of such articles, and the ways to find them.

15. Mr. Dillon's book *Interfaces for Information Retrieval and Online Systems* (New York: Greenwood Press, 1991) contains the following notice:

> "All rights reserved. No portion of this book may be reproduced, by any process or technique, without the express written consent of the publisher."

Lois Chan's books are similarly frozen in non-shifted formats; both her *Guide to Library of Congress Classification* (Englewood, CO: Libraries Unlimited, 1999) and her *Library of Congress Subject Headings* (Libraries Unlimited, 1995) contain identical boilerplate:

> "No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the publisher and the author."

My own books have similar notices. Since the current copyright law protects such works for the life of the author plus seventy years, none of these books is likely to make "the shift" at all. And *should* some of them actually become digital, they will still not be accessible from anywhere, at anytime, by anyone on the Web; their digital versions will likely have physical-place use restrictions not appreciably different from their print counterparts.

16. The figure comes from RLG's Walt Crawford, in an email to me.

17. Lois Mai Chan, "Exploiting LCSH, LCC, and DDC to Retrieve Networked Resources," <http://lcweb.loc.gov/catdir/bibcontrol/chan.html>

18. Lois Mai Chan and Theodora Hodges, "Entering the Millenium: A New Century for *LCSH*," *Cataloging & Classification Quarterly*, 29, 1-2 (2000), 225-34.

19. Regina Reyolds, "Partnerships to Mine Unexploited Sources of Metadata"; and Priscilla Kaplan, "International Metadata Initiatives: Lessons in Bibliographic Control," both available through <http://lcweb.loc.gov/catdir/bibcontrol>.

20. Caplan, Ibid., p. 6.

21. Findable at <http://lcweb.loc.gov/catdir/bibcontrol/thomas.html>

22. There are large commercial and governmental forces at work to get ordinary citizens connected to the Internet *in their homes*. Businesses promote home access because it enables them to target specific audiences and market groups, and to reach them (and their credit cards) immediately and interactively. Government, too, sees civic and educational goals being fostered by the same household hookups to the Net. In remarks made in December of 1999 in the Rose Garden, President Clinton noted the recent successes of public-private partnerships in closing the "digital divide" by wiring all schools and classrooms to the Internet. But he then went on to add, "there's still a lot more to do. We must connect *all of our citizens* to the Internet *not just in schools and libraries, but in homes*, small businesses, and community centers" [emphasis added]. Two months later, in announcing a multi-billion dollar federal program to solve the problem, he said, "Our big goal should be to make connection to the

Internet as common as connection to telephones" (*Washington Post*, 2/3/2000, p. B04). This is a politically popular agenda that will probably be pursued by whoever succeeds Mr. Clinton.

23. Dorothy Gregor and Carol Mandel, "Cataloging Must Change!," *Library Journal* (April 1, 1991), 42-47.

24. Sarah E. Thomas and Jennifer A. Younger, "Cooperative Cataloging: A Vision for the Future," *Cataloging & Classification Quarterly*, 17, 314 (1993), 237-57.

25. Thomas Mann, "'Cataloging Must Change!' and Indexer Consistency Studies: Misreading the Evidence at Our Peril," *Cataloging & Classification Quarterly*, 23, 3-4 (1997), 3-45.

26. Ibid., pp. 37-39.

27. Elaine Svenonius and Dorothy McGarry, "Objectivity in Evaluating Subject Heading Assignment," *Cataloging & Classification Quaterly*, 16, 2 (1993), 5-40.

28. Ann Huthwaite notes in her paper, "At the same time that this revolution has occurred there has been growing pressure on publicly funded institutions to reduce costs. Libraries throughout the world have been cutting back on expenditures and services." ("AACR2 and Its Place in the Digital World," <http://lcweb.loc.gov/catdir/bibcontrol/huthwaite.html>, p. 2.) Is there any doubt that more and more cataloging is being relegated to technicians?